Origin of anti-sickling activity via QSAR modelling

Chuleeporn Phanus-umporn, Saw Simeon, Watshara Shoombuatong and Chanin Nantasenamat

July 7, 2016

Principal Component Analysis

```r
library(readxl)
library(caret)
library(dplyr)
library(ggplot2)
library(cowplot)


df <- read.csv("data.csv")
Activity <- df$Activity
descriptors <- df[, 2:ncol(df)]
data <- descriptors[, which(!apply(descriptors, 2, sd) == 0)]
raw <- cor(data)
df <- data
pca <- prcomp(df, retx=TRUE,scale.=TRUE, center = TRUE)
scores <- pca$x[,1:5]
loadings <- pca$rotation[,1:5]
```

Explained variance, PC scores and loadings

```r
summary(pca)
```

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6
## Standard deviation     2.9993  2.3602 2.23624 1.78266 1.69706 1.57253
## Proportion of Variance 0.1666  0.1032 0.09261 0.05885 0.05333 0.04579
## Cumulative Proportion  0.1666  0.2697 0.36235 0.42120 0.47454 0.52033
##                           PC7     PC8     PC9    PC10    PC11    PC12
## Standard deviation     1.55761 1.52703 1.44023 1.29151 1.2141  1.2007
## Proportion of Variance 0.04493 0.04318 0.03841 0.03089 0.0273  0.0267
## Cumulative Proportion  0.56526 0.60844 0.64685 0.67774 0.7050  0.7317
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     1.16814 1.06793 1.03433 1.03335 1.01087 1.00537
## Proportion of Variance 0.02527 0.02112 0.01981 0.01977 0.01892 0.01872
## Cumulative Proportion  0.75701 0.77813 0.79794 0.81771 0.83664 0.85535
##                           PC19    PC20    PC21    PC22    PC23    PC24
## Standard deviation     0.96591 0.93740 0.91304  0.8664 0.83297 0.81803
## Proportion of Variance 0.01728 0.01627 0.01544  0.0139 0.01285 0.01239
## Cumulative Proportion  0.87263 0.88890 0.90434  0.9182 0.93109 0.94348
##                           PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.72706 0.71224 0.62540 0.61685 0.53243  0.4764
## Proportion of Variance 0.00979 0.00939 0.00724 0.00705 0.00525  0.0042
## Cumulative Proportion  0.95327 0.96267 0.96991 0.97696 0.98221  0.9864
```

```
##                            PC31    PC32    PC33    PC34    PC35    PC36
## Standard deviation      0.42434 0.39393 0.33855 0.30171 0.24283 0.22361
## Proportion of Variance  0.00333 0.00287 0.00212 0.00169 0.00109 0.00093
## Cumulative Proportion   0.98974 0.99262 0.99474 0.99643 0.99752 0.99844
##                            PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation      0.19013 0.14191 0.12601 0.08191 0.06490 0.03102
## Proportion of Variance  0.00067 0.00037 0.00029 0.00012 0.00008 0.00002
## Cumulative Proportion   0.99911 0.99949 0.99978 0.99990 0.99998 1.00000
##                             PC43      PC44      PC45      PC46      PC47
## Standard deviation      1.009e-15 7.972e-16 7.067e-16 5.655e-16 4.832e-16
## Proportion of Variance  0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion   1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC48      PC49      PC50      PC51      PC52
## Standard deviation      4.205e-16 2.225e-16 1.915e-16 1.915e-16 1.915e-16
## Proportion of Variance  0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion   1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC53      PC54
## Standard deviation      1.915e-16 6.011e-17
## Proportion of Variance  0.000e+00 0.000e+00
## Cumulative Proportion   1.000e+00 1.000e+00
scores
```

```
##               PC1         PC2          PC3          PC4          PC5
##  [1,]  0.54208242  0.591083439  0.335994281 -2.390627738  1.877463961
##  [2,]  2.30770772  0.845404745 -0.089501711 -0.657383438 -0.060941371
##  [3,]  2.42121916  0.370960090  0.365209488 -0.878523082 -0.337777727
##  [4,]  0.10057711  0.916006891 -0.161704106 -1.703166002  1.058032186
##  [5,]  0.52304263  0.936187534 -0.177513656 -1.622080602  0.831330409
##  [6,]  0.81808148  0.565018053 -0.100966997 -1.046518196 -0.089408486
##  [7,]  0.57813006  0.859261267  0.172259508 -2.023531128  1.231356541
##  [8,] -3.10639962  0.248554819 -0.351672235 -1.490489629  1.689388879
##  [9,] -2.39002551  2.481061627 -2.204884183 -1.321895553  2.377661446
## [10,] -2.96062430  2.328115910 -2.654807332 -0.044231766 -0.016706061
## [11,] -8.62301768  4.030517069 -5.085805646  0.704737307  4.433023381
## [12,] -9.19361647  3.877571352 -5.535728794  1.982401094  2.038655874
## [13,] -1.23933983 -2.533615686  3.165026340  0.400817329 -0.267617908
## [14,]  2.37322074  0.751698781  0.306177733 -0.827742900 -0.286723866
## [15,]  2.82365724  1.290353625 -0.186044500 -0.921663449 -0.217030958
## [16,]  2.63464294  1.195146911 -0.072024875 -1.326069237  0.119548630
## [17,]  2.61023346  0.466166804  0.251189863 -0.474117294 -0.674357316
## [18,] -2.87800379  0.600179029 -1.312176151  1.597576596 -3.898552681
## [19,] -2.77478403  0.768631653 -1.543922583  1.504577596 -3.833675431
## [20,] -2.67156427  0.937084277 -1.775669015  1.411578595 -3.768798180
## [21,] -3.79082757  1.009307420 -1.262507667 -0.190426791 -0.467623561
## [22,] -2.42756728  1.138833874 -1.804398384  1.503656047 -3.828859773
## [23,] -1.67849995  1.110210637 -1.341555417  0.215000208 -1.601132450
## [24,] -1.65381198  0.499628100 -0.688574068  0.273103742 -2.392105035
## [25,] -4.20939038  2.367733100 -2.796664297  0.350845464 -0.394807819
## [26,] -0.08721945  1.887188093 -0.951791633 -2.629723059  3.481408043
## [27,] -0.51532995  2.046119595 -1.632093941 -0.917311396  0.884095498
## [28,] -4.33681706  1.976382975 -2.704308088  0.845322470 -1.088844936
## [29,] -3.78127988  2.208801598 -2.116361989 -1.361566199  2.202504726
## [30,] -4.38206833  2.218131039 -2.118672974 -1.510998900  2.571300977
## [31,] -6.82725458  3.795051415 -4.889621971  0.620049804  0.134877859
```

```
##  [32,] -4.46783564  2.229982707 -2.823782011  0.397748144 -0.734849162
##  [33,] -3.24315206 -2.170316868  2.456092432  0.803009046 -0.949260422
##  [34,] -2.08427087 -2.573976973  3.196645439  0.238646530  0.185785648
##  [35,] -1.42835413 -2.628822401  3.279045964 -0.003588458  0.068961681
##  [36,] -2.50673640 -2.594157616  3.212454989  0.157561130  0.412487426
##  [37,] -2.55433893 -2.547434984  3.454935031 -0.156987681  0.756501716
##  [38,] -2.92920192 -2.614338259  3.228264539  0.076475730  0.639189204
##  [39,] -4.08808311 -2.210678155  2.487711532  0.640838247 -0.495856866
##  [40,] -3.06156416 -2.583642569  3.283339632 -0.126567714  3.219457648
##  [41,] -0.76197368 -2.462781797  3.073451405  0.698527909 -0.693802593
##  [42,] -1.75875433 -2.596870142  3.229278583  0.232576800 -0.079449059
##  [43,] -2.50673640 -2.594157616  3.212454989  0.157561130  0.412487426
##  [44,] -2.48600406 -2.550672045  3.199047787  0.096184966  0.413864017
##  [45,] -1.42446580 -3.143993742  2.947651997  0.259790977 -0.245720530
##  [46,]  1.63544212 -8.775159193 -6.017361995  0.465911384  0.783524960
##  [47,]  2.26434716 -8.418073302 -6.496664409  0.360998783  0.686577684
##  [48,]  1.27566685 -8.567159474 -6.495438404  0.219903117  0.792991734
##  [49,]  1.86968046 -8.860306301 -6.129634504  0.820486710  0.494406437
##  [50,] -1.73628650 -2.723715655  3.506110542  0.548237136 -0.134672609
##  [51,] -1.99569893 -2.821975495  3.192889123  0.654703814 -0.527046852
##  [52,] -1.21412981 -1.283930410  2.637190535 -0.438788438  0.498158767
##  [53,] -3.58287011 -0.646845256  1.727488263  0.127160175  2.684694803
##  [54,] -1.51174535 -1.499126719  3.195514957 -0.644227549  0.769866916
##  [55,] -0.65771705 -0.363859235  2.681819179  5.294621975  2.662029835
##  [56,] -1.77318910 -1.282287936  2.507644111  0.155958277 -0.123698050
##  [57,] -7.39730107 -1.039803776  1.170246402  0.919105048 -1.928774267
##  [58,] -9.08716316 -1.120526349  1.233484601  0.594763449 -1.021967156
##  [59,] -1.77958286 -0.849298858  1.875618060  0.030915811 -0.931957998
##  [60,] -2.82506569 -1.048124230  2.682086039 -1.229208319  1.875148950
##  [61,] -3.66999674 -1.088485517  2.713705138 -1.391379119  2.328552506
##  [62,] -2.78560915 -0.550681961  1.356251041  1.719330873 -6.948947891
##  [63,]  2.96822365  1.016975461  0.019629543 -0.667933872 -0.388025871
##  [64,]  2.50701370  0.297714181  0.482936295 -0.381118294 -0.739234566
##  [65,]  2.40379394  0.129261557  0.714682727 -0.288119293 -0.804111817
##  [66,]  2.37700992  0.445023644  0.308834713 -0.759089775 -0.332779936
##  [67,]  2.95745021  0.836369025 -0.009285938 -0.475038843 -0.669541658
##  [68,]  3.05105644  0.827210578  0.001829866 -0.939504642 -0.818901558
##  [69,]  2.01378259  0.571189946  0.192377981 -1.422033737  0.415252075
##  [70,]  2.24700613  0.592333107  0.134733131 -1.137061257  0.073674695
##  [71,]  3.52355444  1.649627958  0.182267803  5.082876744  0.963885900
##  [72,]  2.14378637  0.423880483  0.366479563 -1.044062256  0.008797444
##  [73,]  2.24700613  0.592333107  0.134733131 -1.137061257  0.073674695
##  [74,]  2.72422666  0.815225864  0.048358912 -0.760011324 -0.327964278
##  [75,]  1.58102365  0.376808371 -0.009723659 -0.687136827 -0.850055521
##  [76,]  2.24700613  0.592333107  0.134733131 -1.137061257  0.073674695
##  [77,]  3.98654946  1.610080863  0.428597780  5.235886011  1.019296377
##  [78,]  4.60631887  2.171820547  0.246595477  7.534676343  1.668645085
##  [79,]  2.50701370  0.297714181  0.482936295 -0.381118294 -0.739234566
##  [80,]  2.50701370  0.297714181  0.482936295 -0.381118294 -0.739234566
##  [81,]  4.33376621  1.980283084  0.168121979  5.234964462  1.024112035
##  [82,]  1.44969538  1.224892329 -0.637683811 -0.577726741  0.791856303
##  [83,]  1.78022193  0.938555339 -0.114253826 -1.358723648  0.064320287
##  [84,]  3.78010994  1.273175616  0.892090644  5.421884012  0.889541876
##  [85,]  1.58102365  0.376808371 -0.009723659 -0.687136827 -0.850055521
```

```
## [86,]   2.71345322  0.634619428  0.019443431 -0.567116295 -0.609480065
## [87,]   2.61561086  0.406500736  0.876944630 -3.145210290  2.087633227
## [88,]   2.75666581  0.961222028 -0.019601932  0.347168413  0.120227796
## [89,]   2.49100312  0.794082704  0.106003762 -1.044983805  0.013613102
## [90,]   1.01480887  0.247902842 -0.024307204 -0.747147093 -0.970343249
## [91,]   3.29313329 -7.551470540 -6.744024573  0.461669538  0.270984649
## [92,]   2.37322074  0.751698781  0.306177733 -0.827742900 -0.286723866
## [93,]   2.37322074  0.751698781  0.306177733 -0.827742900 -0.286723866
## [94,]   2.47644050  0.920151405  0.074431301 -0.920741900 -0.221846616
## [95,]   1.95075521  0.731518138  0.321987283 -0.908828300 -0.060022088
## [96,]   1.91022571  0.791245876  0.059847756 -0.980752166 -0.342134344
## [97,]   2.11321316  1.046317707 -0.042025431 -1.583685862  0.526185395
## [98,]   2.11321316  1.046317707 -0.042025431 -1.583685862  0.526185395
## [99,]   2.63464294  1.195146911 -0.072024875 -1.326069237  0.119548630
## [100,]  2.91726348  1.281195178 -0.174928696 -1.386129248 -0.366390858
## [101,]  1.48351465  1.588726748 -0.921881475 -0.622831103 -2.254270870
## [102,]  1.87998962  1.025174546  0.015619419 -1.868658343  0.867762774
## [103,]  3.36944720  1.761958399  0.468829659  4.725915938  2.530433052
## [104,]  2.34643671  1.067460868 -0.099670281 -1.298713381  0.184608015
## [105,]  2.11321316  1.046317707 -0.042025431 -1.583685862  0.526185395
## [106,]  3.16866281  1.886811403  0.458513665  5.548123194  3.320202506
## [107,]  2.01344547  0.959698500 -0.171898676 -1.073751167 -0.277257093
## [108,]  0.50278602  1.285675528 -0.518259228 -0.544701783 -0.192898625
## [109,]  1.61182173  0.690664105  0.038925404 -0.956202433  0.191875253
## [110,]  1.26790463  1.627111362 -1.135252575  0.395653548 -7.121763352
## [111,]  1.07551391  0.006214238  0.632194212 -0.669707158 -0.323064000
## [112,]  1.42151851  0.869570404 -0.440122738 -0.282727886 -1.528725845
## [113,] -1.27284127  0.692573595 -0.799360868  0.389633778 -1.616681044
## [114,]  0.51723429  1.221676324 -0.462412145 -0.683681914  0.207204311
## [115,]  1.29836625  1.123331784 -0.431904801 -0.474768428 -1.416284074
```

loadings

```
##                    PC1          PC2          PC3          PC4          PC5
## SubFPC1   -0.05810294 -0.053607816  0.080184731 -0.229260787  0.211118616
## SubFPC2   -0.21839206 -0.034864724  0.051718191 -0.065064904  0.131020359
## SubFPC3   -0.04973999 -0.027711806  0.101456637 -0.071478519  0.114699601
## SubFPC4   -0.05101237  0.026657438  0.022636589 -0.147214513  0.184660909
## SubFPC5   -0.10109808 -0.177325292  0.242276372  0.019042534  0.034065491
## SubFPC10  -0.01580840 -0.025314838  0.060110268 -0.019069912  0.025145836
## SubFPC12  -0.07139294  0.064084716 -0.091977815  0.018354222  0.004405458
## SubFPC14  -0.07139294  0.064084716 -0.091977815  0.018354222  0.004405458
## SubFPC16   0.10877659  0.082328558 -0.004610012 -0.160601238 -0.025158870
## SubFPC17   0.03581775  0.032310460  0.039545496  0.268437667  0.161956642
## SubFPC18  -0.03088638 -0.136710414  0.088395994 -0.063762355 -0.153895614
## SubFPC19   0.01551319  0.026827859 -0.017341350 -0.018436552 -0.073630290
## SubFPC20   0.09631247  0.066769344  0.016423042  0.384221870  0.082760712
## SubFPC23  -0.10955262  0.045377792 -0.053511490  0.028435691  0.195808266
## SubFPC26  -0.12961209  0.040024036 -0.053479372  0.040314021  0.204826314
## SubFPC28   0.04400680  0.034157029  0.005863855  0.123940402  0.079810748
## SubFPC33   0.01515954  0.020684009 -0.011995358 -0.017101408  0.025864066
## SubFPC37   0.03581775  0.032310460  0.039545496  0.268437667  0.161956642
## SubFPC38   0.09631247  0.066769344  0.016423042  0.384221870  0.082760712
## SubFPC48   0.02735157  0.006864330  0.016496051 -0.093102017  0.068187476
## SubFPC49  -0.15896458 -0.192989552  0.298945179  0.061675487  0.015755439
```

4

```
## SubFPC74   -0.12242841 -0.025909267  0.032113813  0.031827004 -0.068450996
## SubFPC84   -0.14750393  0.091087366 -0.132938914  0.083807123 -0.142346106
## SubFPC85   -0.17297174  0.131564706 -0.181123668  0.009048299  0.099385908
## SubFPC88   -0.22405549  0.157658983 -0.215973617  0.048114153 -0.042367615
## SubFPC96    0.09964215  0.073763721  0.037940229  0.476163579  0.167807529
## SubFPC100  -0.09111637  0.009823544 -0.031114865  0.079394157 -0.239772072
## SubFPC101  -0.16442509  0.161270327 -0.202888290 -0.025932333  0.157605576
## SubFPC125  -0.13232229  0.094843271 -0.141903555  0.056493391  0.150129350
## SubFPC135  -0.24063587 -0.092051873  0.165162782  0.071384911  0.064375938
## SubFPC137  -0.26435492 -0.061575821  0.056348784  0.112250931 -0.116254900
## SubFPC139  -0.16401495 -0.183580881  0.296263860  0.062786687  0.020474184
## SubFPC171  -0.10641416 -0.173665814  0.238918409  0.095877089 -0.066885041
## SubFPC172   0.04816850  0.036674207  0.004638664  0.223035505  0.054502244
## SubFPC173   0.08159247  0.067464967 -0.009607087  0.030790655 -0.020084568
## SubFPC174   0.04432609  0.025286528 -0.002312598 -0.048893256 -0.027496292
## SubFPC182   0.01325854  0.027476036 -0.021355036  0.011711822 -0.232615126
## SubFPC183  -0.02912925 -0.009299030  0.025512199  0.050894267 -0.226970528
## SubFPC184  -0.01127183  0.012909831  0.002952536  0.044464640 -0.326410913
## SubFPC188   0.04921847 -0.324215533 -0.273048474  0.031386601  0.045034670
## SubFPC200  -0.01854233 -0.021653213  0.047170851  0.004616553 -0.004040297
## SubFPC274  -0.13331755  0.064527482 -0.092321360  0.115579176 -0.365706031
## SubFPC275  -0.17236190  0.125988512 -0.135362228 -0.035921394 -0.024571659
## SubFPC287  -0.24077942 -0.218820079  0.001385368  0.129691052 -0.109714201
## SubFPC295  -0.26311675  0.085533393 -0.151860338  0.089510232 -0.046895789
## SubFPC296   0.04921847 -0.324215533 -0.273048474  0.031386601  0.045034670
## SubFPC297   0.04921847 -0.324215533 -0.273048474  0.031386601  0.045034670
## SubFPC298   0.04921847 -0.324215533 -0.273048474  0.031386601  0.045034670
## SubFPC299   0.04921847 -0.324215533 -0.273048474  0.031386601  0.045034670
## SubFPC300  -0.26461254  0.043789873 -0.026888975 -0.101420235  0.210853312
## SubFPC301  -0.26461254  0.043789873 -0.026888975 -0.101420235  0.210853312
## SubFPC302  -0.27062424 -0.068485442  0.007728233  0.060939877  0.016394132
## SubFPC303  -0.07791090 -0.298057462  0.116115055  0.035206997  0.050559370
## SubFPC307  -0.26041809  0.020053409 -0.123320337  0.094828413 -0.205098789
```

PCA plot with clusters

```r
km <- kmeans(scores, center=5, nstart=5)
ggdata <- data.frame(scores, Cluster=km$cluster)
### paper numbering
library(grid)
set.seed(23)
x <- ggplot(ggdata, aes(x = PC1, y = PC2, colour = Cluster)) +
  geom_point(aes(fill=factor(Cluster)), size=5, shape=20, pch = 21, alpha = 0.8) +
  ggtitle(" ") +
  stat_ellipse(aes(fill=factor(Cluster)), colour = "black",
              geom="polygon", level=0.95, alpha=0.2) +
  guides(color=guide_legend("Cluster"),fill=guide_legend("Cluster")) +
  #geom_text(aes(label=compoundnumber), size=7, hjust=0.5, vjust= 1.5, alpha=0.45) +
  theme(
    legend.position=("none"),
    #plot.title = element_text(size=20, face="bold", colour="black", vjust = 2, hjust=-0.07),
    panel.border = element_rect(linetype = "solid", colour = "black", fill = NA, size = 1),
```

```
    axis.text.y = element_text(size = 15),
    axis.ticks.length = unit(0.3, "cm"),
    axis.text.x = element_text(size = 15),
    legend.title=element_blank(),
    axis.title.x = element_text(color="black", size=20),
    axis.title.y = element_text(color="black", size=20)) +
  coord_cartesian(ylim = c(-15, 15), xlim = c(-15, 15))
x
```